Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 45 (2015) 790 – 799

Procedia Computer Science

International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)

Application of Fuzzy-Rough Oscillation on the Field of Data Mining (Special Attention to the Crime Against Women at Tripura)

Susmita Roy [a,*], Sharmistha Bhattacharya [b]

[a]*National Institute of Technology, Agartala, Barjala, Jirania, Tripura, Pin-799046, India*
[b]*Tripura University(A Central University),Suryamaninagar,Agartala-799022,Tripura,India*

Abstract

The aim of this present paper is to show the application of fuzzy-rough oscillation concept on the field of data mining especially in the subject of crime against women at Tripura. The concept of fuzzy-rough oscillatory region was first introduced in the year of 2009[1]. In this present investigation we have collected crime against women data from Tripura (North Eastern State of India). After that the set of attributes are reduced by the aid of network techniques such as fuzzy cognitive map (FCM). The collected qualitative data are then converted into quantitative data by fuzzy membership value lying between [0, 1] using technique "Direct methods with multiple object". Lastly we apply "Fuzzy-Rough Oscillatory Region Algorithm" on the said data.

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).

Keywords: Fuzzy set; Rough set; Fuzzy cognitive map; Fuzzy oscillation; Topology; Rough topology.

1. Introduction

Rough Set theory was introduced in 1982 by Z.Pawlak[2]. It is a new mathematical approach to data analysis and data mining. It is another mathematical approach to vagueness. Various Researchers worked a lot to solve various data mining and decision making problem by the tools of Rough Set and Fuzzy Set theory.

* Corresponding author. Tel.: +91-9436565145.
E-mail address: rsusmita@yahoo.com.

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).
doi:10.1016/j.procs.2015.03.155

Some of the techniques of data reduction introduced earlier such as Boolean reasoning rule extraction etc. But this method doesn't work properly for all type of data. So a new method of data reduction using Fuzzy Cognitive Map is applied in this paper. After data reduction, by the method of Fuzzy-rough Oscillation introduced by the authors, some decision making problems such as Crime against Women at Tripura are solved. Since the data set is too large for hand computation hence an algorithm is introduced so that the data set may be handled by the help of Computer.

The objective of the present paper is

- To collect the data about Crime against Women at Tripura by preparing a questionnaire. In this questionnaire all possible attributes are kept. Then data are collected by the method of cluster sampling.
- To reduce data with the help of fuzzy cognitive map.
- To convert the qualitative data to quantitative by fuzzy membership grade.
- To use 'Fuzzy-rough oscillatory region algorithm' to the above data.
- To develop a C-programming for the above algorithm.

The motivation behind applying the application of fuzzy-rough oscillation concept in the study of crime against women at Tripura is given below:

- No such mathematical approach is yet been applied to the data of Crime against women at Tripura.
- The factors which are responsible for occurring crime against women may be linguistic terms which can be translated into fuzzy membership grades.
- The number of attributes for crime against women are large that can be reduced with the help of data reduction method available in rough set theory.
- Last but not the least, the women bearing the same attributes, some of them may be secured or some of them may not be which is the actual situation for using rough set theory.

Nomenclature

$\in$	Belongs to
N	Set of Natural numbers
Σ	Summation
μ_α	membership function of a fuzzy subset α
$\subseteq$	Sub set
$\cap$	Intersection
$\neq$	Not equal to
ϕ	Phi
I^x	The set of all fuzzy subsets on X
Λ	Big lambda
$\rightarrow$	Arrow indicating direction

2. Preliminaries

In this section we shall give those notions, which are necessary to define concepts used in this paper. The concept of Fuzzy Set was introduced by Professor L.A Zadeh in 1965[3]. It is proved to be very useful in all the branches of Mathematics. The concept of fuzzy set is defined as below:

Let X be an ordinary set. A fuzzy subset α in X is the collection of ordered pairs $(x, \mu_\alpha(x))$ with $x \in X$ and a membership function $\mu_\alpha : X \rightarrow [0, 1]$. The value $\mu_\alpha(x)$ of x denotes the degree to which an element x may be a member of α.Thus a fuzzy subset α of X is denoted by $\alpha = \{(x, \mu_\alpha(x)) : x \in X\}$,where $\mu_\alpha(x) = 1$ indicates strictly the containments of the elements x in α and $\mu_\alpha(x) = 0$,denotes x does not belong to α.

2.1. Direct Method with Multiple Experts

The concept of "Direct methods with Multiple Experts" was introduced to find the membership value of the attribute set.

Assuming that n objects are asked for some $x \in X$ to evaluate the proposition "x belongs to A" as either true or false, where A is a fuzzy set on X that represent a linguistic term associated with a given linguistic variable. Given a particular element $x \in X$, let $a_i(x)$ denote the answer of object x_i ($i \in N$). Let $a_i(x) = 1$ when the proposition is valued by expert I is true, and $a_i(x) = 0$ when it is valued as false. Then $A(x) = \Sigma a_i(x)/n$, $i = 1, 2, \ldots, n$, is the membership value of each attribute.

2.2. Fuzzy Cognitive Map

Fuzzy Cognitive Map (FCM) was introduced by Bart Kosko in the year 1986[4]. FCMs have a major role to play mainly when the data concerned is an unsupervised one. Further this method is most simple and an effective one as it can analyze the data by directed graph and connection matrices.

An FCM is a directed graph with concepts like policies, events etc. as nodes and causalities as edges. It represents causal relationship between concepts. If the increase (or decrease) in one concept leads to the increase (or decrease) in another, then we give value 1. If there exists no relation between two concepts the value 0 is given. If increase (or decrease) in one concept decreases (or increases) another, then we give the value -1. Thus FCMs are described in this way. When the nodes of the FCM are fuzzy sets then they are called as fuzzy nodes. FCMs with edge weights or causalities from the set {-1,0,1} are called simple FCMs.

2.3. Rough set

Let U be a finite non-empty set, called universe and R be an equivalence relation on U, called indiscernibility relation. By R(x) we mean that the set of all y such that xRy ,i.e. $R(x) = [x]_R$ is containing the element x. Let X be a subset of U .We want to characterize the set X w.r.t R. The Lower approximation of a set X w.r.t R is the set of all objects, which surely belong to X i.e. $R_*(X) = \{x: R(x) \subseteq X\}$.The Upper approximation of X w.r.t R is the set of all objects, which are partially belonging to X. i.e. $R^*(x) = \{x: R(x) \cap X \neq \phi\}$.

The difference of the upper and the lower approximation is a boundary region. It consists of all elements that cannot be classified uniquely to the set or its complement, by employing available knowledge. Thus any rough set, in contrast to a crisp set, has a non-empty boundary region. The lower approximation is called interior and the upper approximation is called closure of the set.

2.4. Fuzzy-rough Oscillation

In this section the concept of fuzzy-rough oscillation is described. The above defined lower approximation set forms the topology. The elements of this topology are open sets i.e. the elements of lower approximation are open sets. And the elements of upper approximations are closed sets. With this topology we can now introduce some operators as follows:

The operator Λ, Int,Cl and V : $I^x \to I^x$ is defined as

(i) $\Lambda_{a_j}(x) = \inf\{\mu_{a_j}(x_i) : \mu_{a_j}(x_i) \geq \mu_{a_j}(x), x_i \in G$, G is an open set , $j=1,2,\ldots,n\}$
$= \hat{I}$,otherwise.

(ii) $\mathrm{Int}_{a_j}(x) = \mathrm{Sup}\{\mu_{a_j}(x_i): \mu_{a_j}(x_i) \leq \mu_{a_j}(x), x_i \in G$,G is an open set, $j = 1,2,\ldots,n\}$
$= \phi$,otherwise

(iii) $\mathrm{Cl}_{a_j}(x) = \inf\{\mu_{a_j}(x_i): \mu_{a_j}(x_i) \geq \mu_{a_j}(x), x_i \in G$,G is an closed set, $j = 1,2,\ldots,n\}$
$= \hat{I}$,otherwise.

(iv) $V_{a_j}(x) = \sup\{\mu_{a_j}(x_i): \mu_{a_j}(x_i) \leq \mu_{a_j}(x), x_i \in G$,G is an closed set, $j = 1,2,\ldots,n\}$
$= \phi$,otherwise

where $\mu_{a_j}(x_i)$is the membership value of any particular attribute a_j of any object x_i and $\mu_{a_j}(x)$ is the membership value of unknown object for a particular attribute a_j

An operator $O^o: I^x \to I^x$ such that $O^o{}_{a_j}(x_i) = \Lambda_{a_j}(x_i) - Int_{a_j}(x_i)$

This operator is said to be fuzzy rough open oscillatory operator.

and an operator $O^c: I^x \to I^x$ such that

$$O^c{}_{a_j}(x_i) = Cl_{a_j}(x_i) - V_{a_j}(x_i)$$

This operator is said to be fuzzy rough closed oscillatory operator.

We have, $h_{a_j}(x_i) = \inf\{\Lambda_{a_j}(x_i), Cl_{a_j}(x_i)\} - \sup\{Int_{a_j}(x_i), V_{a_j}(x_i)\}$

From the above relation following cases may arise:

Case I: $h_{a_j}(x_i) = \Lambda_{a_j}(x_i) - Int_{a_j}(x_i)$, the object may lie in boundary region.

We need to check the numerical value of $h_{a_j}(x_i)$

Hence three cases may arise:

i) If $h_{a_j}(x_i) \geq 0.5$, then the attribute tends to go towards the boundary region.

ii) If $h_{a_j}(x_i) < 0.5$,then the object tends towards lower approximation.

iii) If $h_{a_j}(x_i) = 0$, then the object is definitely in the lower approximation

Case II: $h_{a_j}(x_i) = \Lambda_{a_j}(x_i) - V_{a_j}(x_i)$

Case III: $h_{a_j}(x_i) = Cl_{a_j}(x_i) - Int_{a_j}(x_i)$,

Case IV**:** $h_{a_j}(x_i) = Cl_{a_j}(x_i) - V_{a_j}(x_i)$,

In the above case II, case III, case IV, the object may lie in boundary region. So we need to check the numerical value of $h_{a_j}(x_i)$.Here two cases may arise:

i) If $h_{a_j}(x_i) \geq 0.5$, then the attribute tends to go outside region.

ii) If $h_{a_j}(x_i) < 0.5$, then the object tends to go towards the lower approximation.

CaseV: $h_{a_j}(x_i) = \hat{I} - V_{a_j}(x_i)$,

In this the attributes lie on outside region.

We need to check the numerical value of d which is computed by the difference between the membership value of the attribute a_j and $V_{a_j}(x_i)$ of the object.

Let this difference be $d = V_{a_j}(x_i) - \mu_{a_j}(x_i)$.Three cases may arise:

i) If $d \geq 0.5$, the attribute lie in outside region.

ii) If $d < 0.5$,the attribute lie in boundary region .

iii) If $d = 0$,the attribute lies completely in the boundary region.

CaseVI: $h_{a_j}(x_i) = \hat{I} - Int_{a_j}(x_i)$, the attribute may lie in outside or boundary.

In this case we need to check the numerical value of d ,the difference between the membership value of the attribute a_j and $Int_{a_j}(x_i)$ of the object .Let this difference be $d = Int_{a_j}(x_i) - \mu_{a_j}(x_i)$.

Three cases may arise:

i) If $d \geq 0.5$, the attribute lie in outside region.

ii) If $d < 0.5$,the attribute lie in boundary region.

iii) If $d = 0$, the attribute is in the lower approximation

CaseVII: $h_{a_j}(x_i) = \Lambda_{a_j}(x_i) - \phi$,

In this case, we need to find the difference between $\Lambda_{a_j}(x_i)$ and the membership value of the attribute a_j of the unknown object. Let $d = \Lambda_{a_j}(x_i) - \mu_{a_j}(x_i)$.

Hence three cases may arise:

i) If $d \geq 0.5$, then the attribute tends to go outside region.

ii) If $0 < d < 0.5$, then the attribute tends to go lower approximation.

iii) If $d = 0$, then the attribute must lie in L.A.

CaseVIII: $h_{a_j}(x_i) = cl_{a_j}(x_i) - \phi$,

In this case the attribute lie on Upper approximation.
Let $d = cl_{a_j}(x_i) - \mu_{a_j}(x_i)$.
Hence three cases may arise:
i) If $d \geq 0.5$, then the attribute tends to go outside region.
ii) If $0 < d < 0.5$, then the attribute tends to go lower approximation.
iii) If $d = 0$, then the attribute must lie completely in the boundary region.
Case IX: $h_{a_j}(x_i) = \hat{I} - \phi$,This is an unstable case. In this case we cannot make any decision because our dataset does not permit to deal with such type of object.
So while drawing any decision by the help of height of oscillation at first we have to check the membership values of the height of oscillation and then the structure.
Finally we can draw a conclusion about an unknown object or about the pattern of that unknown object with the help of value of height of oscillation as follows:
Let us distribute these cases in three parts:
Stable: In this case we can make perfect decision about an unknown object.
Case (a): If the height of oscillation of each attribute lies on the lower approximation then the object lie on lower approximation and full decision can be drawn in favour of lower approximation.
Case (b): If the height of oscillation of each attribute lies outside, then the object is in outside region.
Unstable: In this case we cannot make any decision because our dataset does not permits to deal with such type of object.
Oscillating: In this case we can make decision with some error.
A part of developed C-program for the above algorithm is given below:

```
#include<stdio.h>
#include<conio.h>
#include<math.h>
void main()
{
int i,j,k,m,n,r,c,z,s,l=0,bo=0,o=0,us=0;
float d,b[50],low[50][50],upper[50][50],p[50],lemda[50],Int[50],cl[50],v[50],h[50];
clrscr();
printf("Give the no. of object and attribute of lower approx.:");
scanf("%d %d",&m,&n);
printf("\n Give the membership value of lower approximation:");
for(i=1;i<=m;i++)
for(j=1;j<=n;j++)
scanf("%f",&low[i][j]);
printf("\nGive the no. of object and attribute of upper approx.:");
scanf("%d%d",&r,&s);
printf("\n Give the membership value of upper approximation:");
for(i=1;i<=r;i++)
for(j=1;j<=s;j++)
… … … … … … … …
```

Fig.1. A part of C-programming developed

3. Data Collection and Data Reduction

3.1. Data Collection

To achieve our target i.e. to study about crime against women at Tripura we have collected required data at primary stage. In connection to the collected data we at first have prepared a questionnaire so that it becomes very easy to collect data from different kind of people. The process of data collection is such that one questionnaire is given to a person (male/ female). The response of that person will be answered by either yes or no or by tick mark. Then the filled in questionnaire was recorded and counted separately. Data are collected by collecting sample from every race, cast, region etc.

3.2. Data Reduction

Depending on questionnaire data are collected from a sample of 500 persons by the method of cluster sampling with each cluster of different tribe, race, religion etc. which is too large to be mentioned in this paper. Next we have drawn a network diagram with the help of required attributes such as type of crime, reason of crime, attitude of police, police station, steps for security, women cell, and duties of women cell, which is shown below:

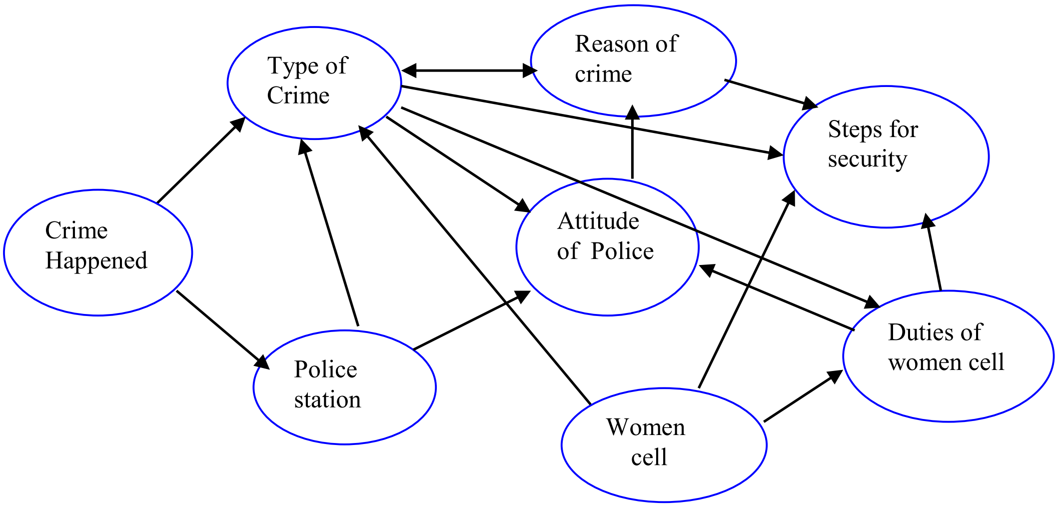


Fig.2. Diagram of relationships between the attributes

Table 1. Different types of crime

Sl no.	Types of crime(2)	Responses	
		Number	Percentage
1	Dowry Harassment	100	20%
2	Family discord	200	40%
3	Kidnapping	15	3%
4	Rape	50	10%
5	Harassment at the work place	25	5%
6	Female feticide	25	5%
7	Eve-teasing	25	5%
8	Abandonment of the girl child	25	5%
9	Any other	35	7%
	Total	500	100%

After collecting data we need to reduce the number of attributes. Now with the help of Fuzzy Cognitive Map different relations are shown below:

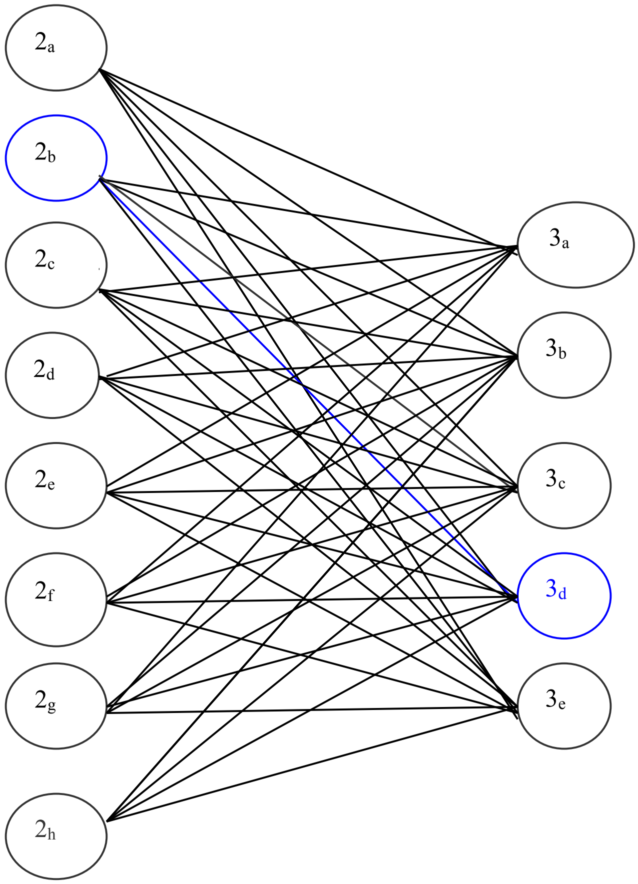
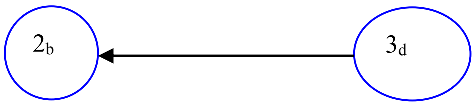


Fig. 3. Relation between type of crime and reason of crime

Table 2. The responses of the people

	3_a			3_b			3_c			3_d			3_e		
	dir	inver	nil	dir	inver	nil	dir	inver	nil	dir	inver	nil	dir	inver	Nil
2_a (100)			0		+40				0		+60				0
2_b (200)		+20			+10		+10	+40		**+40**	**+50**		+1 0	+10	
2_c (15)		+10			+5				0			0			0
2_d (50)		+20				0		+20				0		+10	
2_e (25)			0			0		+25				0			0
2_f (25)		+5			+10				0		+10				0
2_g (25)		+5				0		+20				0			0
2_h (25)			0			0			0			0		+25	

In the above table ‘dir’ means directly related and ‘inver’ means inversely related and ‘ nil’ means no relation. From the above data, we conclude that 2_b and 3_d are inversely related with value +10 i.e.,



Proceeding in the same way with the help of fuzzy cognitive map we get the main attributes as follows:

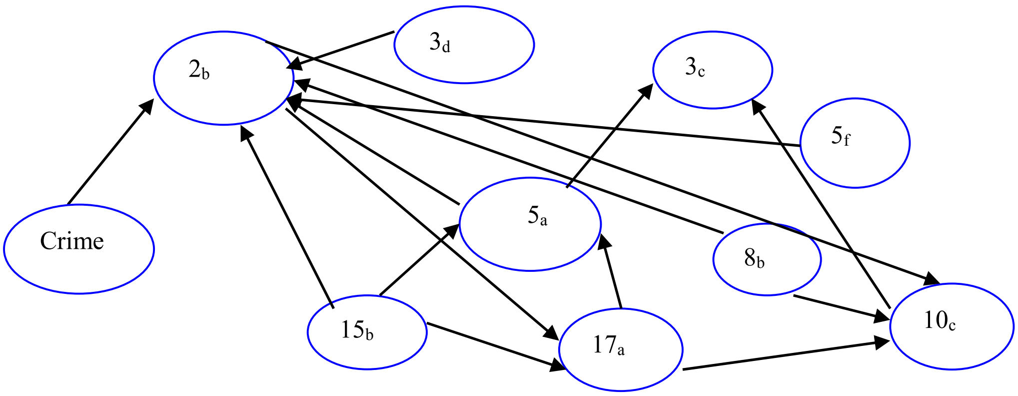


Fig. 4. Diagram of relationships between the attributes after data reduction

Where	2_b	-	family discord/domestic violence (type of crime)
	3_d	-	economic pressure (reason of crime)
	3_c	-	increase of women participation in the work place (reason of crime)
	5_a	-	women’s awareness of their rights (steps for security)
	5_f	-	social protests (steps for security)
	10_c	-	police undecided (attitude of police)
	8_b	-	no enough police station
	15_b	-	no enough women cell
	17_a	-	awareness of women (duties of women cells)

By the help of Fuzzy Cognitive Map we can conclude that, the maximum crime is occurring due to domestic violence/family discord. The main reason of such type of crime is either for economic pressure or increase of women participation in the work place. The necessary steps required for their security is women’s awareness of their rights and social protests. There are not enough police stations in this state and also not enough women cell. Most of the cases the attitude of polices are undecided. Duties of the women cells should make awareness program for the women. By this method we have reduced the attributes of the data set and then we will apply fuzzy rough oscillation theory for decision making related to this problem.

4. Application of Fuzzy-Rough Oscillation on Crime Against Women at Tripura

In this section an example is taken to show the application of fuzzy-rough oscillation. The data set is taken from crime against women at Tripura where attributes consist of linguistic variables. At first we transferred it by fuzzy membership value between [0,1]. “Direct Method with Multiple Object” to find the membership value of the attribute as follows:

First of all we make a questionnaire .The questionnaire contains various types of questions and corresponding to every question there are various answers marking a_i, i =1,2,….n.

Suppose n objects are asked for questions and they replied by a_i, $i\in N$
Then we valuate the membership of the answer a_i by

$\mu(a_i)$=(total no of a_i)/n.

The values are converted to fuzzy values by "Direct methods with multiple object" and we get the following table 3:

Table.3. Data converted to fuzzy values

Object	Type of crime (Family Discord)	Reason of crime		Steps for security		Police station (No)	Attitude of police (Un decided)	Wom-en cell (No)	Duties of women cell (Awareness of women)	Secured or not
		women participation	Economic Pressure	Women's awareness	Social protests.					
x_1	0.6	0.0	0.8	0.5	0.3	0.7	0.5	0.8	0.5	0
x_2	0.7	0.7	0.1	0.4	0.6	0.8	0.6	0.7	0.3	1
x_3	0.8	0.2	0.8	0.3	0.2	0.6	0.7	0.6	0.4	0
x_4	0.9	0.6	0.3	0.5	0.7	0.5	0.8	0.5	0.5	0
x_5	0.2	0.4	0.5	0.9	0.8	0.8	0.6	0.7	0.9	1
x_6	0.7	0.7	0.1	0.4	0.6	0.8	0.6	0.7	0.3	0
x_7	0.6	0.0	0.8	0.5	0.3	0.7	0.5	0.8	0.5	1
x_8	1	0.1	0.7	0.7	0.3	0.5	0.6	0.6	0.7	0
x_9	0.3	0.8	0.0	0.9	0.8	0.8	0.6	0.7	0.9	1
x_{10}	0.2	0.5	0.5	0.9	0.8	0.8	0.6	0.7	0.9	1

From the above Table 3,
Let X={x_1,x_2,x_3,x_4,x_5,x_6 x_7, x_8, x_9, x_{10}}. The lower approximation and upper approximations are Lower Approximation(i.e. open set) ={x_3,x_4,x_8} and Upper Approximation(i.e. closed set) ={x_1,x_2,x_3,x_4,x_6,x_7,x_8} respectively ,whereas the boundary line cases are the objects ={ x_1,x_2,x_6, x_7}
Let us find the decision of an unknown woman (P) whose attributes are as follows:

2_b=0.85,3_c=0.4,3_d=0.7,5_a=0.3,5_f=0.4,8_b=0.8,10_c=0.6,15_b=0.5,17_a=0.3[by Simple Random Sampling Without Replacement(SRSWOR)].Then using fuzzy-rough oscillating algorithm we get,

h_{2b} (P) = inf{0.9,0.9}-sup{0.8,0.8}= 0.1<0.5[Λ-Int],by case I, the attribute tends to go toward the lower approximation.
h_{3c} (P) = inf{0.6,0.6}-sup{0.2,0.2}= 0.4<0.5[Λ-Int],by case I, the attribute tends to go toward the lower approximation.
h_{3d} (P) = inf{0.7,0.7}-sup{0.7,0.7}= 0.0<0.5[Λ-Int],by case I, the attribute is in the lower approximation.
h_{5a} (P) = inf{0.3,0.3}-sup{0.3,0.3}= 0.0<0.5[Λ-Int],by case I, the attribute is in the lower approximation.
h_{5f} (P) = inf{0.7,0.4}-sup{0.3,0.4}= 0.0<0.5[Cl-V],by case IV, the attribute tends to go toward the lower approximation.
h_{8b} (P) = inf{I,0.8}-sup{0.6,0.8}= 0.0<0.5[Cl-V],by case IV, the attribute tends to go toward the lower approximation.
h_{10c} (P) = inf{0.6,0.6}-sup{0.6,0.6}= 0.0<0.5[Λ-Int],by case I, the attribute is in the lower approximation.
h_{15b} (P) = inf{0.5,0.5}-sup{0.5,0.5}= 0.0<0.5[Λ-Int],by case I, the attribute is in the lower approximation.
h_{17a} (P) = inf{0.4,0.3}-sup{ϕ,0.3}= 0.0<0.5 [Cl-V],by case IV, the attribute tends to go toward the lower approximation. i.e. by Stable case (a) the object is not secured.

5. Conclusion

The concept studied in this paper helps us to draw conclusion about the pattern of an unknown object. Stochastic methods had also been used in various data mining technique in various papers in the literature. But stochastic process deals with only time dependent data. The method described in this paper helps us to deal with all type of data set. A computer programming is required for this purpose. This program will help to draw conclusion about an unknown object easily. Though there is lots of software as Rosetta, KDD, LERS, LEM1, LEM2, MLEM2 etc but they are costly, not available and not so easy to deal. But this program can be done by any mathematical programming language like C, MATHEMATICA, MATLAB software which is available everywhere and also decision can be drawn easily which is not possible by the other software available in market.

It helps not only in Social and Human sciences, also for the NGO's so that they can study the coping strategies of victims as well as health and other related facilities available to take care of their physical and mental trauma and suffering. There are many researchers in Mathematics who wish to extend their research into the area of Social Science, especially that of Economics. I hope this work may provide a helping hand for them.

Acknowledgements

In the last of this paper, I would like to acknowledge some person's credit other than those whom this paper might not be completed. The Chairperson, Tripura Women Commission, Tripura, The HO.D, Mathematics, Tripura University, my fellow research scholars and students of Tripura University, Tripura ,India.

References

1. Bhattacharya (Halder) S , Roy S. Some More Results On Fuzzy-rough Oscillatory region and it's Application in Decision Making. *Proceeding of International Conference on Rough Sets, Fuzzy Sets and Soft Computing*. Serial Publication;ISBN 978-81-8387-417-5. 2011; p.206-216
2. Pawlak Z. Rough sets. *International Journal of computer and Information Sciences* 1982;**11**:341-356.
3. Zimmerman HJ. Fuzzy set theory and its Application. Kluwer Academic Publisher 1991.
4. Kosko B. Fuzzy Cognitive Maps. *International Journal of Man-Machine Studies* 1986;**24:**65-75.
5. Klir George J, Yuan Bo. *Fuzzy sets and Fuzzy Logic*, p. 280-285.
6. Mukherjee A , Halder S. Some More Results On Oscillatory Region Of A Fuzzy Set. *International conference on Modeling and Simulation*. CIT, Coimbatore 2007;p.665-670.